AWARD NUMBER: W81XWH-15-2-0016

TITLE: An Interactive Visualization Framework to Support Exploration and Analysis of TBI/PTSD Clinical Data

PRINCIPAL INVESTIGATOR: Dr. Jesus Caban

CONTRACTING ORGANIZATION: The Geneva Foundation
Tacoma, WA 98402

REPORT DATE: May 2017

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE ***(DD-MM-YYYY)*** May 2017	**2. REPORT TYPE** ANNUAL	**3. DATES COVERED** ***(From - To)*** 15 Apr 2016 - 14 Apr 2017

4. TITLE AND SUBTITLE An Interactive Visualization Framework to Support Exploration and Analysis of TBI/PTSD Clinical Data	**5a. CONTRACT NUMBER**
	5b. GRANT NUMBER W81XWH-15-2-0016
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Dr. Jesus Caban jmolstre@genevausa.org	**5d. PROJECT NUMBER**
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) The Geneva Foundation 917 Pacific Ave, Suite 600 Tacoma, WA 98402	**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Gay Hayden Grants Specialist US Army Medical Research Acquisition Activity Phone: 301-619-9883	**10. SPONSOR/MONITOR'S ACRONYM(S)** USAMRAA
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

We propose to design, develop, and validate an interactive visualization framework that physicians assessing TBI/PTSD patients with comorbid symptoms can use to explore and analyze clinical data and that researchers can use to hypothesize new research questions. The primary aims of this project are to (1) extend our interactive visual analytic framework which combines multiple clinical measurements to allow it to be used to explore large collections of clinical data and (2) validate the effectiveness of such visualization systems among clinicians that treat service members.

15. SUBJECT TERMS
Data Visualization, Health Information Technologies, TBI/PTSD, Open Source Tools

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT U	**b. ABSTRACT** U	**c. THIS PAGE** U	UU	16	**19b. TELEPHONE NUMBER** ***(include area code)***

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

Table of Contents

Page

1. INTRODUCTION: An Interactive Visualization Framework to Support Exploration and Analysis of TBI/PTSD Clinical Data

2. **KEYWORDS:** Data Visualization, Health Information Technologies, TBI/PTSD, Open Source Tools

3. **ACCOMPLISHMENTS:** The overarching goals of this project are to (1) address the gap between the acquisition of clinical measurements and the diagnosis step by providing an institutive, flexible, and customizable interactive data visualization framework and (2) validate the system among clinicians treating service members diagnosed with TBI / PTSD.

During the first year of the award, a significant amount of work was accomplished. The work accomplished during the first year and continued during the second year can be summarized by describing four unique tools that have been developed to support the first goal of the award.

1. We developed and extended our **VISXplore** software application to be more modular, flexible, and scalable by modifying the code and improving the libraries that were used. VisXplore is a clinical data visualization system that has been designed to perform group or single-subject analysis of multivariate tabular, hierarchical, or temporal clinical data. The visualization system has been developed as a robust, flexible, and easy-to-use application that can be used by either researchers or clinicians. The framework has been tested with different clinical datasets including a dataset of patients diagnosed with post-traumatic stress disorder (PTSD) and complex comorbid conditions. See attached paper for more information about VISXplore.

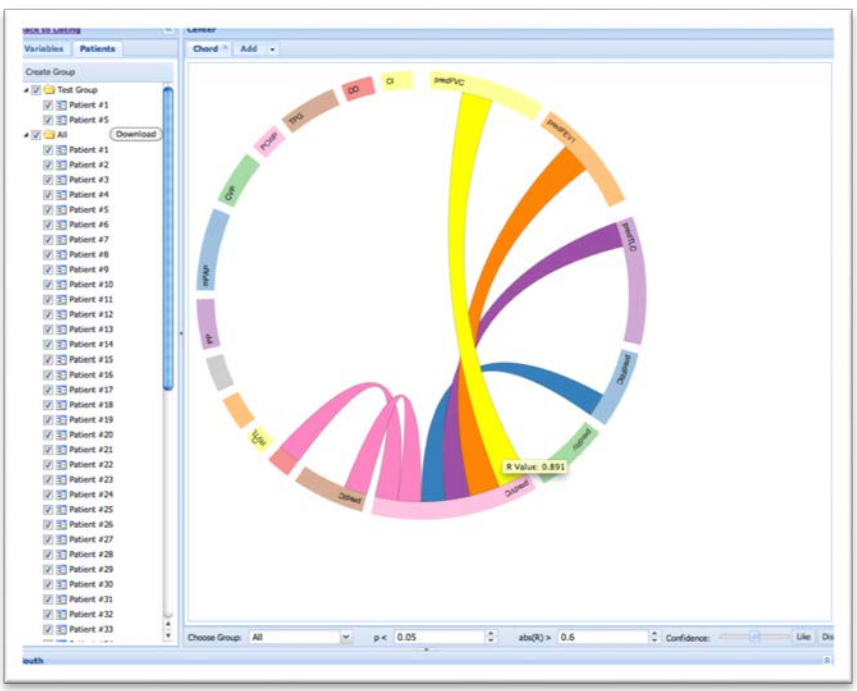


Figure 1: Screenshot of our VisXplore tool designed to perform group and single-subject analysis of multivariate tabular, hierarchical, and temporal clinical data.

2. We developed **CoFlow**, an interactive multi-view and exploratory visualization tool designed to analyze longitudinal EHR data. The primary visualization technique adopts a human-in-the-loop approach and multiple synchronized views can be constructed for the user to explore longitudinal clinical sequences. The exploration tasks are aided by context information visualized using pixel-

oriented and histogram techniques. The context can inform further exploration as well as help analyze diagnoses patterns across various temporal and categorical dimensions. CoFlow facilitates two broad use cases: assessing relative proportions of patients from the overall population which exhibit a certain sequence of diagnoses and identifying cohorts of patients that exhibit a certain combination of diagnoses. See attached draft paper for more information about CoFlow.

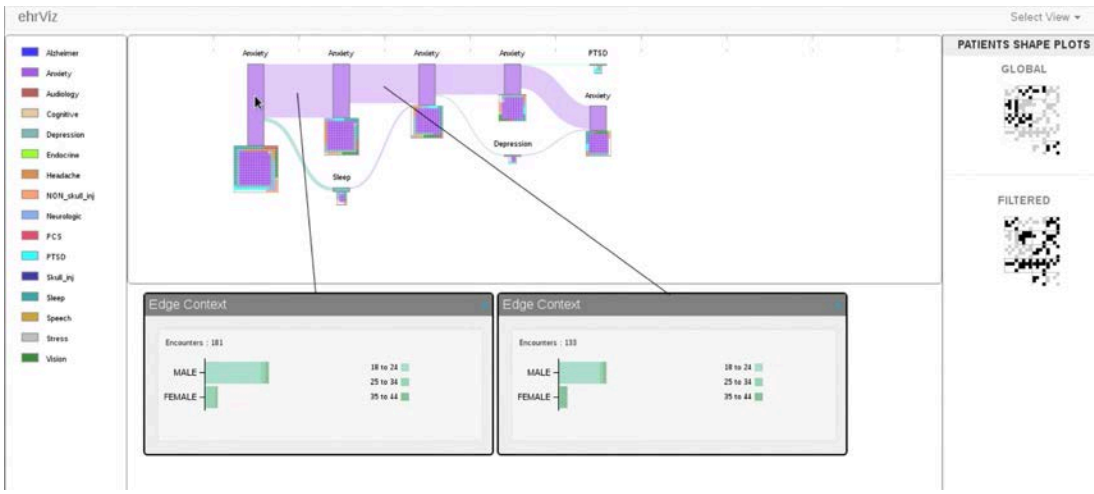


Figure 2: CoFlow, an interactive multi-view and exploratory visualization tool designed to analyze longitudinal EHR data.

3. We designed and developed a graph-based visualization method to interactively analyze the longitudinal clinical trajectory of a group of patients. The system allows users to select a specific set of events or conditions, filter the data based on different thresholds, and compare different cohorts while using an interactive virtual space that expands as the user continues to analyze and explore the data. To build our graph, we first determine all the unique clinical diagnosis on our dataset, treat them as temporal events, and create a node for each possible diagnosis. Then we draw a path/force with an arrow from each node to every other node, representing a patient moving from one diagnosis to the next within their clinical trajectory. At each node we store the number of times that the diagnosis occurred in our dataset and at each path we stored the number of times that unique transition (i.e. Anxiety ->Depression) occurred in our dataset. See attached paper for more information about this technique.

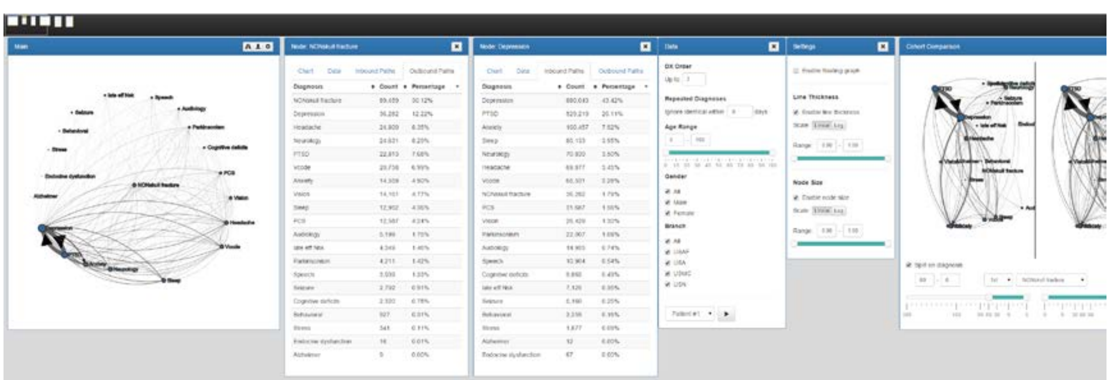


Figure 3: A scalable framework has been designed to visually explore large collections of temporal sequences by creating a graph-based visualization method.

4. We designed and developed a system to visually summarize a collection of temporal sequences using Adaptive Frequency Mining and Graph-based Event Modeling. Summarizing a collection of temporal sequences is a difficult task given the irregular and variable patterns often found in longitudinal events. Across a wide array of domains, researchers and analysts seek to determine ways to identify common temporal paths, to build trajectories between individual events, and to understand the relationships between different events. While these tasks continue to be difficult on small and structured datasets, they are increased tenfold on temporal sequences that are noisy, irregular, and voluminous in size. Approaches that enable analysis of temporal sequences of large or small, noisy or clean, irregular or structured datasets open new opportunities to identify key information embedded within longitudinal data. During the last year a scalable framework has been designed to visually explore large collections of temporal sequences by combining advanced event mining algorithms with visualization techniques to overcome some of the challenges and complexities of the data. Our approach uses a novel adaptive window-based frequency sequence mining algorithm for identifying common subsequences to build an overview of the common longitudinal trajectories, a graph-based event model to illustrate the relationship between events, and a modified sankey diagram to display common paths while providing multiple levels of details. The different modeling and visualization techniques have been integrated within an easy-to-use and scalable interactive graphical system that promotes exploration of large, irregular, and complex datasets. The system has been tested with a comprehensive clinical dataset of 98,342 patients and 8.7 million longitudinal events showing the effectiveness of the techniques within large and complex datasets. See attached draft paper for more information about this tool.

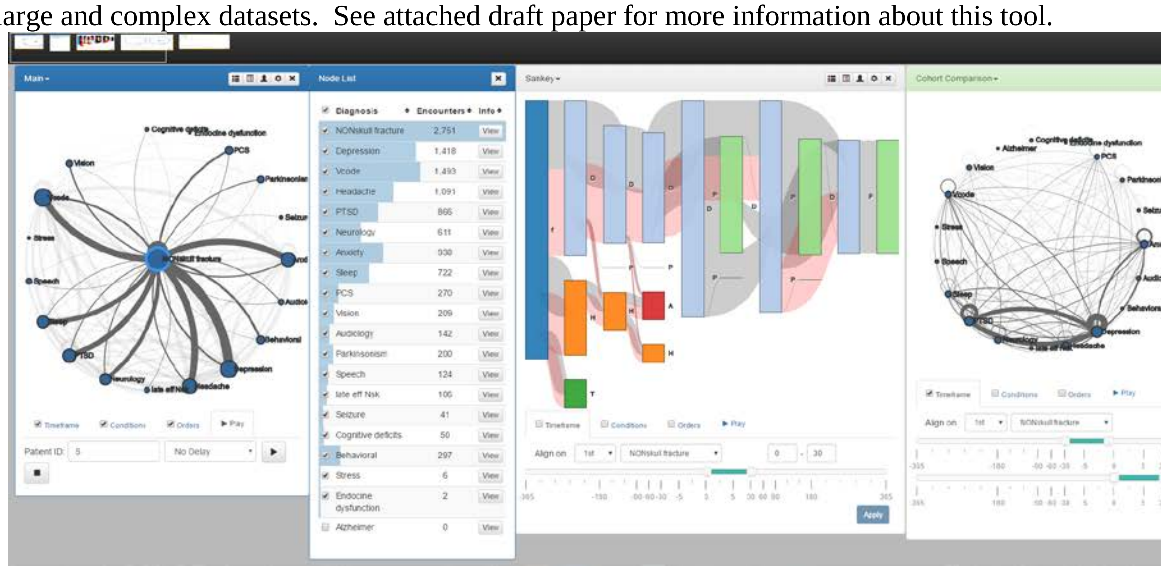


Figure 4: Scalable framework has been designed to visually explore large collections of temporal sequences by combining advanced event mining algorithms with visualization techniques to overcome some of the challenges and complexities of the data.

During the second year of the award, we expanded on and evaluated our previous work as well as accomplished a significant amount of work on developing a new system. The work can be summarized by describing the expansion of our work and the new system that we have developed.

5. We expanded on our work to visually summarize a collection of temporal sequences by building a novel visualization and system that enabled for better analysis and exploration of the generated summaries. Our new visualization, named event summary diagram, removes the visual

complexity and analytical complexity faced by Sankey diagrams. Furthermore, through an interactive interface, the user is able to modify the inputs into our algorithm that we have developed to be able to modify the length of time that they are analyzing and how the time should be grouped together. Through an evaluation of our algorithm and visualization we have found that the presented information is easier to understand, because of the reduced complexity, and the presence of time in our approach. See attached draft paper for more information about this tool.

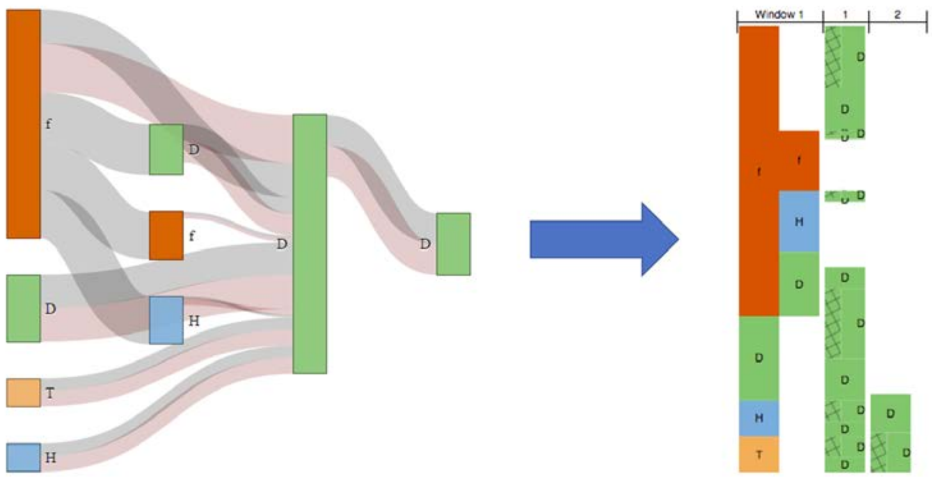


Figure 5: An example of our novel visualization method, event summary diagram, that reduces the visual complexity from a Sankey diagram and the amount of analytical complexity required by consolidating the edges between nodes and encoding important information as textures.

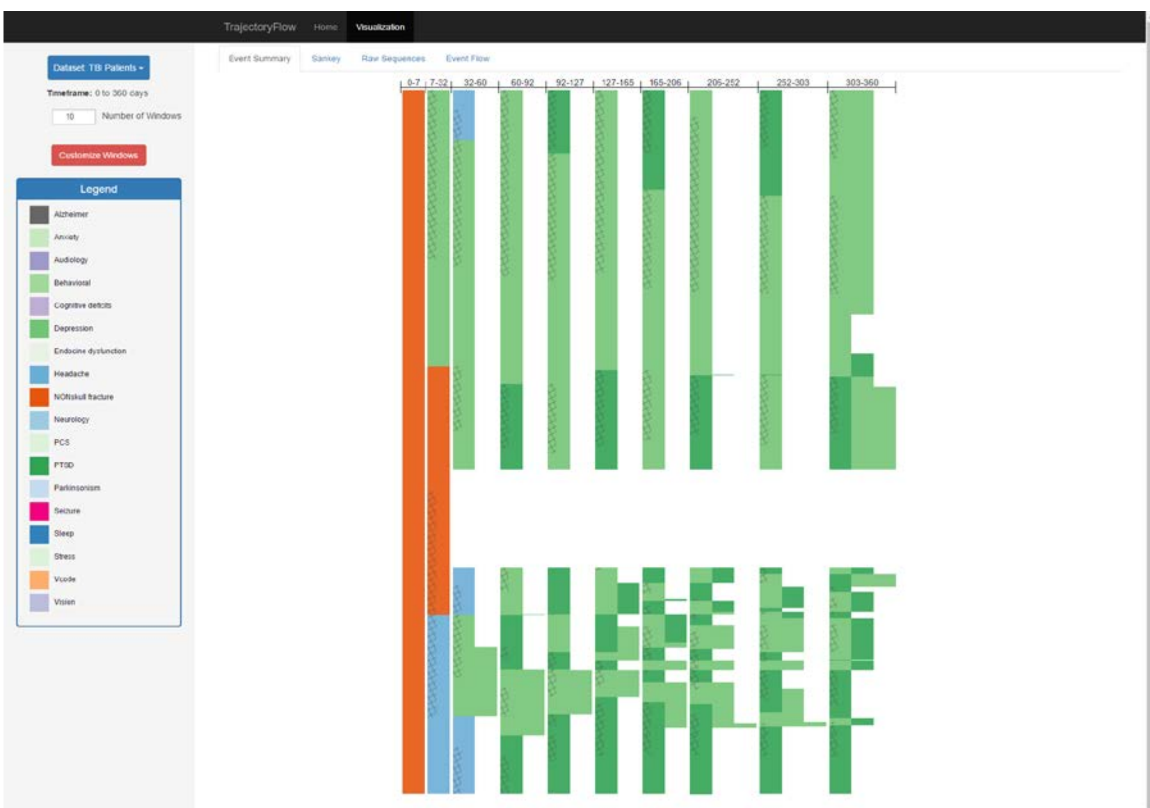


Figure 6: Interactive system has been designed to allow users to customize the time frames and time grouping methods input into our algorithm (through the "Customize Windows" feature). Furthermore, users are able to analyze the events that occurred in a longitudinal dataset of patients to understand the trajectories that patients take and common events that occur in each span of time.

6. We designed and developed a system that assists providers in analyzing relevant patient information by summarizing the various elements of an Electronic Health Record system. Electronic Health Records (EHRs) contain a significant amount of longitudinal information about a patient including pre-existing conditions, earlier diagnosis, previous treatments, active medications, base-line measurements for different clinical results, and much more. Unfortunately, data integration within an EHR and across different EHRs continue to be a limiting factor that threatens patient safety and the efficiency of healthcare providers. The disparate nature of the clinical data even within a single EHR often results in clinicians having to access and review a number of reports, modules, and tabs to access different data elements and clinical results. Due to the fragmented nature of EHR interfaces and the number of interactions that are needed to access clinical data, clinicians often spend a considerable part of their time going through the EHR of a patient in order to get a comprehensive overview and to be able to provide quality care. Data visualization and the integration of analytic models within graphical interfaces present a unique opportunity to effectively combine multiple clinical data sources and reduce the cognitive burden that disparate reports often have for end-users. With the ability of visualization techniques to summarize different data elements, we present a timeline-based framework to effectively aggregate and summarize the disparate clinical data of a patient enclosed within an EHR. The interface combines a set of visualization techniques with machine learning summarization approaches to optimize the process of understanding a patient's history through views that allow for easily skimming and jumping through time, filters for limiting the amount of information shown, and a hierarchy of summaries that provide an interface to view and compare different time frames. See attached draft paper for more information about this tool.

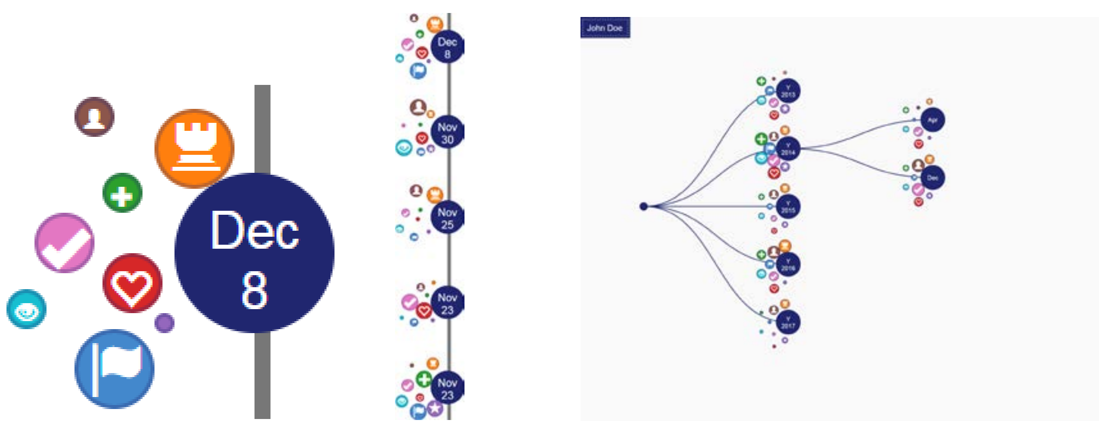


Figure 7: (Left) We represent each date in a patient's timeline as a summary node, where a set of attribute nodes are attached on the left hand side of the node. Each attribute node encodes the significance and volume of a data element present in the EHR. (Middle) A longitudinal timeline of a patient's history which contains a series of summary nodes. This allows for a cliniciain to easily and quickly be able to see the type of information and sate of the patient over time. (Right) A tree summarizing the patient's EHR. First the clinician is able to see information broken down by year and by interactively expanding/collapsing nodes they are able to subdivide a node into smaller time frames.

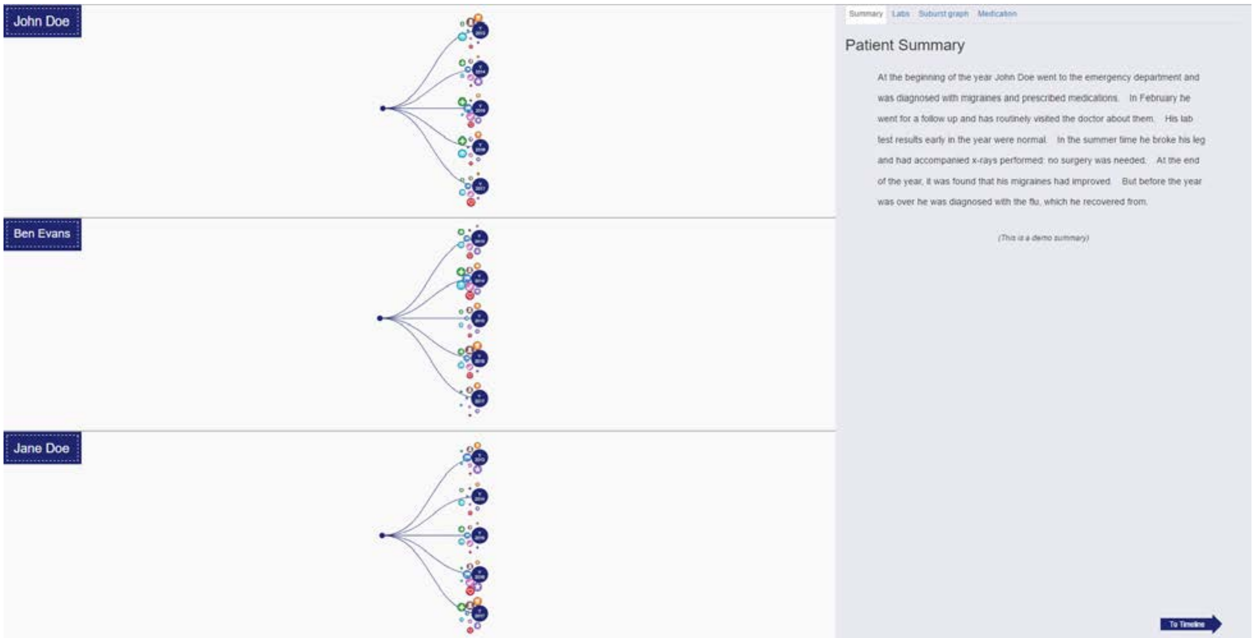


Figure 8: The Summary View which combines both the summary tree and a summary panel into a single display to present the patient(s) that a clinician is attempting to analyze and understand. In the summary panel clinicians are able to view various aspects of a patient for the selected node in the tree: an automatically generated textual summary and a history of lab values, diagnoses, and medications. Through this summary view, comparison of patients at multiple time scales and levels of detail are easily accessible.

Figure 9: The entire patient timeline view with a quick navigation menu on the right and the summary nodes on the left. In the middle is a visualization of the patient's lab test results. The different colored sections of each bar indicate whether the patient is in the normal range.

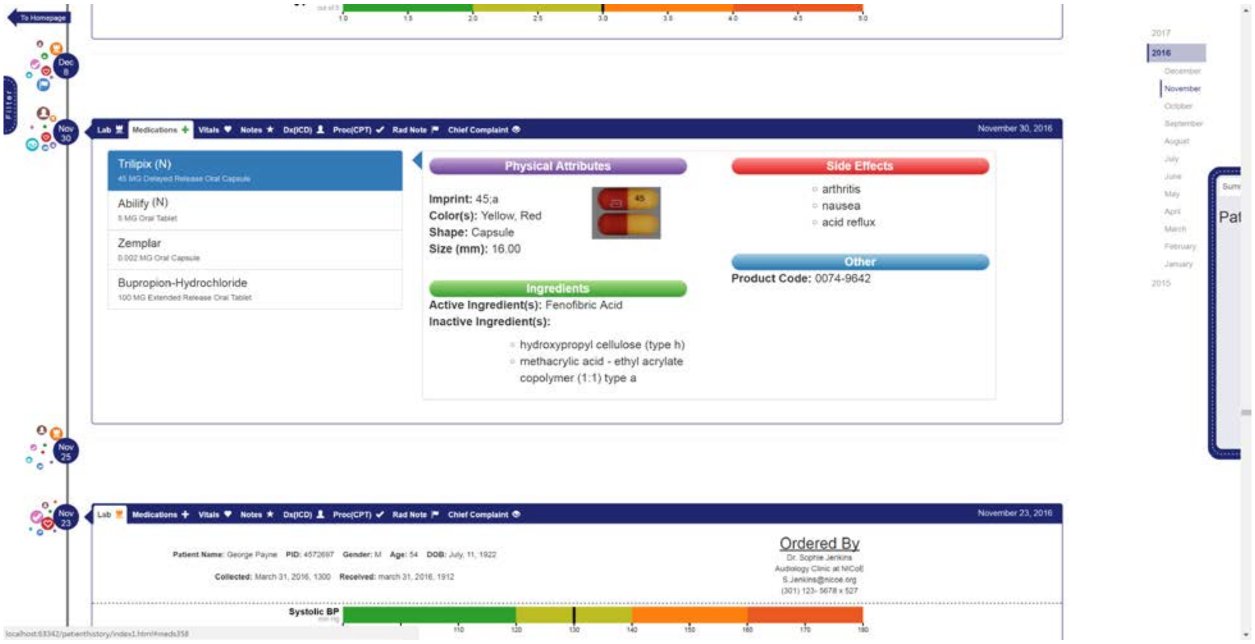


Figure 10: A list of the patient's medications and a variety of information related to the medication. Clinicians are able to identify medications that were newly prescribed on this date through the symbol "new" that appears next to the name of a medication.

What were the major goals of the project?

The first major goal of this project was to design a visual analytic framework that combines multiple clinical measurements and allows the exploration of large collections of clinical data. In addition, the second major goal (year 2) is to validate the effectiveness and usability of different visualization techniques for exploring large collections of clinical variables with complex associations.

The first major goal of the second year was to iterate on our work for summarizing large, longitudinal datasets by expanding on our previous work and developing a novel visualization method for analyzing a large dataset of events. In addition, we validated the effectiveness of our visualization techniques throughout the year. The second major goal of the second year was to build a system that summarizes the various elements of a patient's EHR data. Through these two goals, we have addressed the many difficulties associated with clinical variables and data from both a big data and a singular patient-provider perspective.

What was accomplished under these goals?

During the first year, four different systems were prototyped and developed to perform visualization of tabular, hierarchical, and longitudinal data. First, **VisXplore** was enhanced to become a clinical data visualization system to perform group or single-subject analysis of multivariate tabular, hierarchical, or temporal clinical data. Second, **CoFlow** was developed as an interactive multi-view and exploratory visualization tool designed to analyze longitudinal EHR data. Third, a **graph-based visualization**

technique was developed to visually explore the frequency of patients going from one specific clinical diagnosis to other diagnosis. Finally, a **visual summarization approach** was created and tested with thousands of mTBI patients. Each of the tool has a corresponding draft paper describing the design and techniques. See attachments.

During the second year, two different systems were extensively prototyped and developed to effectively summarize the various data elements that are present in Electronic Health Records (EHRs). First, a novel visualization method, **event summary diagrams**, and a corresponding system were built to enable for a large dataset of events to easily be understood through a top-down interactive exploration. This visualization was evaluated with a dataset of thousands of mTBI patients and shown to reduce the visual complexity and analytical capacity required compared to existing techniques. Second, a **timeline-based framework** for aggregating and summarizing EHRs was extensively researched, designed, and developed to overcome the challenges that exist in EHR systems where data integration is lacking and the disparate nature of data creates difficulties for clinicians. Through this framework, a clinician is able to view the entire history of a patient at multiple time scales and develop an understanding of the patient state over time. Each of these tools have a corresponding draft paper describing the design and techniques. See attachments.

What opportunities for training and professional development has the project provided?

"Nothing to Report."

How were the results disseminated to communities of interest?

The first prototype visualization tool that used a graph-based illustration to show clinical data resulted in a paper publication in the 2015 Workshop on Visual Analytics in Healthcare (see attached paper). Three additional papers describing the other systems are currently in draft mode. See attached documents.

What do you plan to do during the next reporting period to accomplish the goals?

Continue to enhance the different system and start the validation process.

4. IMPACT:

What was the impact on the development of the principal discipline(s) of the project?

The four prototype systems that have been designed have generated great interest among multiple providers, researchers, and administrators. Two senior individuals at the Defense Health Agency (DHA) have seen the systems and are interested in looking into how we can integrate some of those tools within the DHA enterprise enclave. In addition, widely recognized researchers from Johns Hopkins University (JHU) are interested in how to used our visualization techniques for population health.

What was the impact on other disciplines?

The impact of our work is touching multiple disciplines and research domains including clinical informatics, health IT, computer science, medicine, and population health.

What was the impact on technology transfer?

"Nothing to Report."

What was the impact on society beyond science and technology?

"Nothing to Report."

5. CHANGES/PROBLEMS:

Changes in approach and reasons for change

"Nothing to Report"

Actual or anticipated problems or delays and actions or plans to resolve them

The project and actions are a little bit behind schedule due to the challenges of finding qualified candidates that can obtain the credentials needed to work within a DoD facility.

Changes that had a significant impact on expenditures

No changes on expenditure.

Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents

"Nothing to Report"

6. PRODUCTS:

- **Publications, conference papers, and presentations**

  Journal publications. Nothing to Report

  Books or other non-periodical, one-time publications. Nothing to Report

  Other publications, conference papers, and presentations

  - Filip Dabek and Jesus J Caban "VisXplore: Flexible Visualization System for Analyzing Complex Clinical Datasets", Workshop on Visual Analytics in Healthcare, ACM Digital Library, Oct. 2014

- Filip Dabek, J. Chen, A. Garbarino, and Jesus J. Caban, "Visualization of Longitudinal Clinical Trajectories using a Graph-based Approach", Workshop on Visual Analytics in Healthcare, ACM Digital Library, Oct. 2015
- Akshay Peshav, Jian Chen, and Jesus J. Caban, "CoFlow: Interactive Visual Exploration of Temporal Encounters in Electronic Health Records" [Draft]
- Filip Dabek, Jian Chen, and Jesus Caban, "Visual Summarization of a Collection of Temporal Sequences using Adaptive Frequency Mining and Graph-based Event Modeling" [Draft]

- **Website(s) or other Internet site(s)**

"Nothing to Report"

- **Technologies or techniques**

The design and development of our different visualization tools have produced novel techniques including:

- Novel graph-based approach to visualize clinical trajectories
- New pixel-based visualization method that works as a look-ahead tool for patients
- Novel sequence modeling algorithm to summarize longitudinal trajectories

1. **Inventions, patent applications, and/or licenses**

"Nothing to Report"

2. **Other Products**

- The four different software tools

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS
What individuals have worked on the project?

Name: Jesus Caban, PhD
Project Role: PI
Contribution to Project: Dr. Caban has organized meetings, tracked progress of the project, and evaluated various visualization techniques for exploring large clinical data.

Name: Jian Chen, PhD
Project Role: Co-PI
Contribution to Project: Dr. Chen has helped assist in meetings and researched existing visualization techniques in order to identify a new technique to deploy for the visual analytics framework.

Name: Elizabeth Jimenez
Project Role: Developer
Contribution to Project: Ms. Jimenez has begun implementing an interface for the visual analytics framework, in addition to developing and evaluating a visualization technique.

Name: Filip Dabek

Project Role: N/A (Data Scientist for the National Intrepid Center of Excellence)
Contribution to Project: Mr. Dabek has begun implementing an interface for the visual analytics framework, in addition to developing and evaluating a visualization technique.

Name: Jiaoling Chen
Project Role: Jr. Software Developer
Contribution to Project: developed an interface for the visual analytics framework.

Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?

"Nothing to Report."

What other organizations were involved as partners?

University of Maryland Baltimore County
1000 Hilltop Cir, Baltimore, MD 21250

Partner's Contribution: Collaboration.

8. SPECIAL REPORTING REQUIREMENTS:
See attachments.

- W81XWH-15-2-0016 Year 2 QuadChart.ppt

8. APPENDICES:
Grant Tracker

An Interactive Visualization Framework to Support Exploration and Analysis of TBI/PTSD Clinical Data

DMRDP

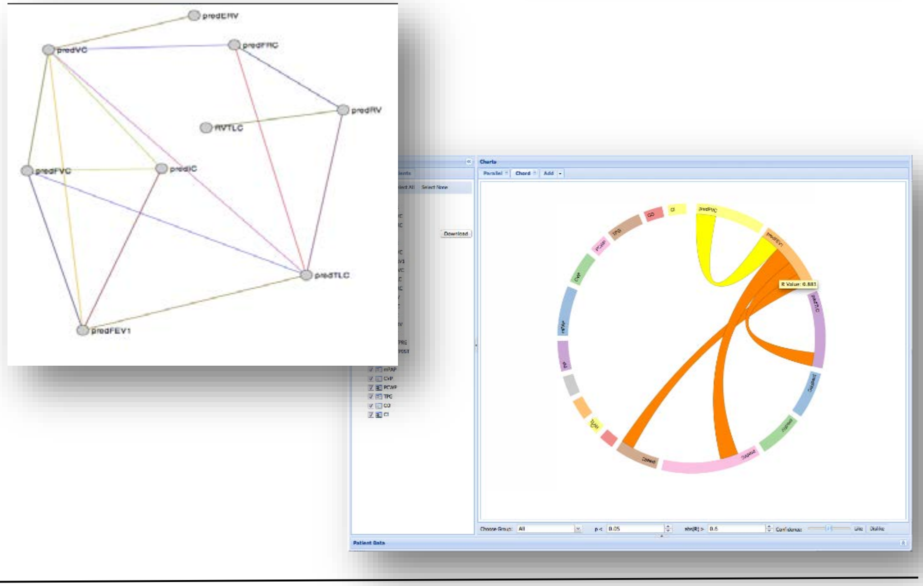
PI: Dr. Jesus J. Caban **Org:** National Intrepid Center of Excellence (NICoE) / Geneva

Problem, Hypothesis and Military Relevance

- **Problem**: The large number of evaluation techniques to assess TBI patients poses challenges to clinicians who must integrate disparate measurements to understand the patient's condition.

- **Hypothesis**: In order to successfully analyze a large number of multi-modal clinical variables and integrate different evaluation protocols, new visual analytics techniques and applications are needed.

- **Military Relevance**: The computational models and visual analytical interfaces will use TBI data of military personnel.



Proposed Solution

- **Objective 1**: The design and development of an intuitive framework to visually analyze and explore a large number of clinical variables.
- **Objective 2:** Perform a usability study to validate application with clinical and research staff treating service members diagnosed with TBI/PTSD
- **Objective 3:** Establish a research community. The software application will be shared with researchers and clinicians at different DoD and federal organizations as a validated software application to visually integrate and analyze large number of clinical variables.

Budget Expenditure to Date:

Projected Expenditure: $452K

Actual Expenditure: $244K

Updated: 5 July 2017

Timeline and Cost

Activities	FY15	FY16	FY17
R&D visual analytics application			
Compare application with existing techniques			
Perform a usability study			
Validate software application with TBI Data			
Software Integration with other tools and libraries			
Share application			
Estimated Total Budget ($1.016)	$229K	$223K	

Appendix A

Task	Timeline	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
Task 1.1	1-3																																				
Task 1.2	1-3																																				
Task 1.3	3-18																																				
Task 1.4	9-12																																				
Task 1.5	9-15																																				
Task 1.6	12																																				
Task 1.7	15-16																																				
Task 1.8	20-24																																				
Task 1.9	23-24																																				
Task 1.10	24																																				
Task 2.1	4-8																																				
Task 2.2	9-15																																				
Task 2.3	16-19																																				
Task 2.4	20-22																																				
Task 2.5	22-24																																				

Task	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	Original SOW
Task 1.1																																					100%
Task 1.2																																					100%
Task 1.3																																					90%
Task 1.4																																					100%
Task 1.5																																					100%
Task 1.6																																					80%
Task 1.7																																					80%
Task 1.8																																					80%
Task 1.9																																					80%
Task 1.10																																					80%
Task 2.1																																					100%
Task 2.2																																					90%
Task 2.3																																					80%
Task 2.4																																					80%
Task 2.5																																					80%
									On Track				Plan Ahead																								88%